

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-139278
(43)Date of publication of application : 20.05.1994

(51)Int.Cl. G06F 15/40

(21)Application number : 04-287261
(22)Date of filing : 26.10.1992

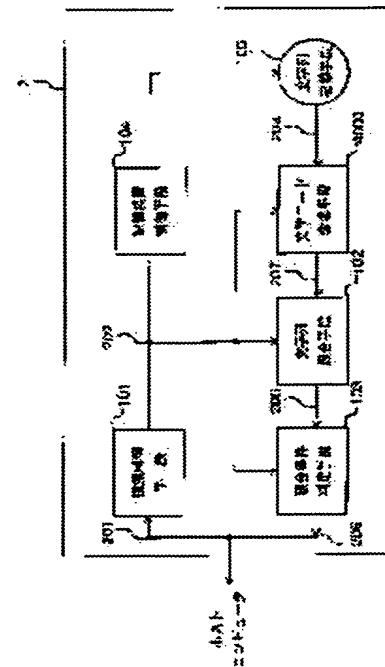
(71)Applicant : HITACHI LTD
(72)Inventor : KAWAGUCHI HISAMITSU
TADA KATSUMI
KATO KANJI

(54) CHARACTER STRING RETRIEVING DEVICE QUIPPED WITH CHARACTER CODE CONVERTING FUNCTION

(57)Abstract:

PURPOSE: To provide a compact character string retrieving device in which the state transition table of a little memory amount can be realized by decreasing an unnecessary slot.

CONSTITUTION: This device is equipped with a character string storage means 105 which stores a text, character code converting means 4000 which fetches each character from the text read from the character string storage means, outputs a code corresponding to the character when a character code is the character code included in a retrieval term, and outputs a specific code when the character code is not the character code included in the retrieval term, retrieval control means 101 which converts the character code of the preliminarily applied retrieval term into the corresponding code, and character string collating means 102 which collates whether or not the code string corresponding to the plural retrieval terms transmitted from the retrieval control means is present in the code string outputted from the character code converting means in a batch.



BEST AVAILABLE COPY

LEGAL STATUS

[Date of request for examination] 31.03.1999

[Date of sending the examiner's decision of rejection] 29.10.2002

[Kind of final disposal of application other than
the examiner's decision of rejection or
application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

[of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

H06-139278

[0027]

[Effect] The principle of matching in the present invention employing the above means is explained.

[0028] In the search control means 101, first, the character codes designated to the search term are found, and a serial code, for example, is assigned to these character codes. Such a character code correspondence table is set for the character code converter means 4000. In addition, an automaton generated from the search term, which is converted into the serial number, is set for the letter string matching means 102.

[0029] In the search, the text read out from the letter string storing means 102 is loaded letter-by-letter into the character code conversion means 4000, and it is determined whether or not the character code is included in the search term. If the character code is included, in the serial number corresponding to the letter is output, and if not, a given number, for example 0(zero), is output.

[0030] Whether or not a code string corresponding to a plurality of search terms is present among the code strings output from the character code conversion means 4000 is found collectively by the matching in the letter string matching means 102.

[0031] The above principle is explained using a specific example.

[0032] First, the setting information generation method is explained.

[0033] In the present example, two terms "CAT" and "DOG" are employed as search terms as in the automaton in Fig. 4. There are six types of letters used in these search terms. The letters "C", "A", "T", "D", "O", and "G" are used in the search terms corresponding to "(1)", "(2)", "(3)", "(4)", "(5)" and "(6)", respectively, and the letters not used in the search terms correspond to "(0)". This character code conversion table is set for the character code conversion means 4000. The character

code conversion table of the present example set in the above manner is shown in Fig. 7.

[0034] In the same way, the search terms are converted to "(1) (2) (3)" corresponding to "CAT" and "(4) (5) (6)" corresponding to "DOG". The automaton generated from these converted search terms in the above manner is shown in Fig. 8. The letters assigned to the transition of the automaton of Fig. 4 are converted into the corresponding serial numbers. This automaton is set to the letter string matching means 102.

[0035] In the present example, types of the character code can be reduced to 7 types from "(0)" to "(6)", which is approximately one-fortieth of the conventional 256 types of this code conversion. Therefore, the state transition table requires approximately one-fortieth of the capacity.

[0036] The matching operation is explained next. In the following description, "HOTDOG" is used as an example of text.

[0037] First, when "H" is input to the character code conversion means 4000, the corresponding "(0)" is output by the character code conversion table shown in Fig. 7.

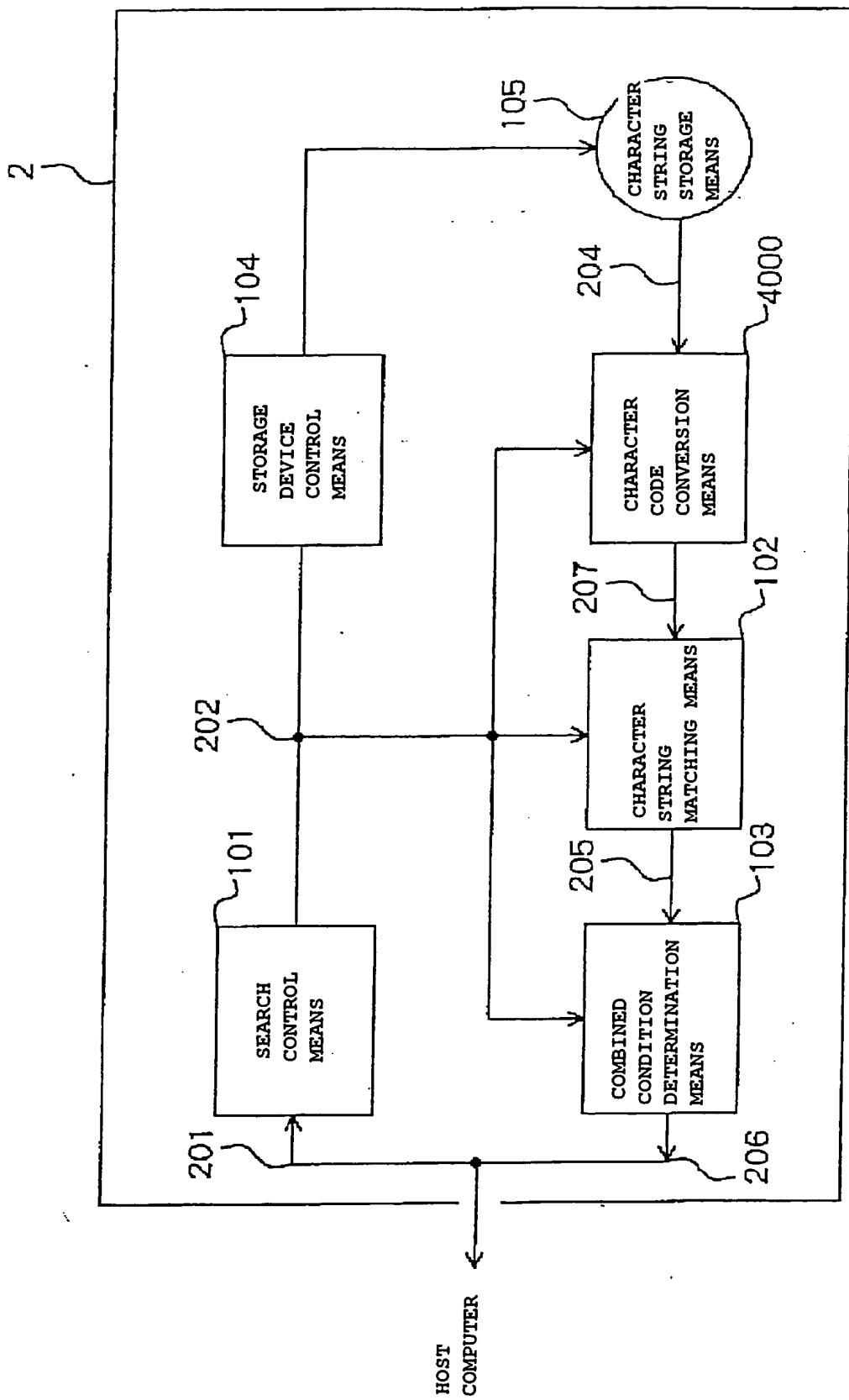
[0038] Next, when "O" is input, the corresponding "(5)" is output by the character code conversion table.

[0039] In the same manner, "T", "D", "O", and "G" are input one after another, "(3)", "(4)", "(5)" and "(6)" are obtained in series as outputs. In other words, "(0) (5) (3) (4) (5) (6)" is transmitted to the letter string matching means 102 as output text.

[0040] In the letter string matching means 102, "(1) (2) (3)" corresponding to "CAT" or "(4) (5) (6)" corresponding to "DOG" are searched in the output texts transmitted from the character code conversion means 4000 based on the automaton shown in Fig. 8. Here, the input of "(0) (5) (3) (4) (5) (6)", "(4) (5) (6)" corresponding to "DOG" is matched.

[0041] By the above operations, the state transition table can be implemented in a memory capacity several tenths that of the prior art, and it is possible to provide a compact letter string search apparatus.

Fig. 1



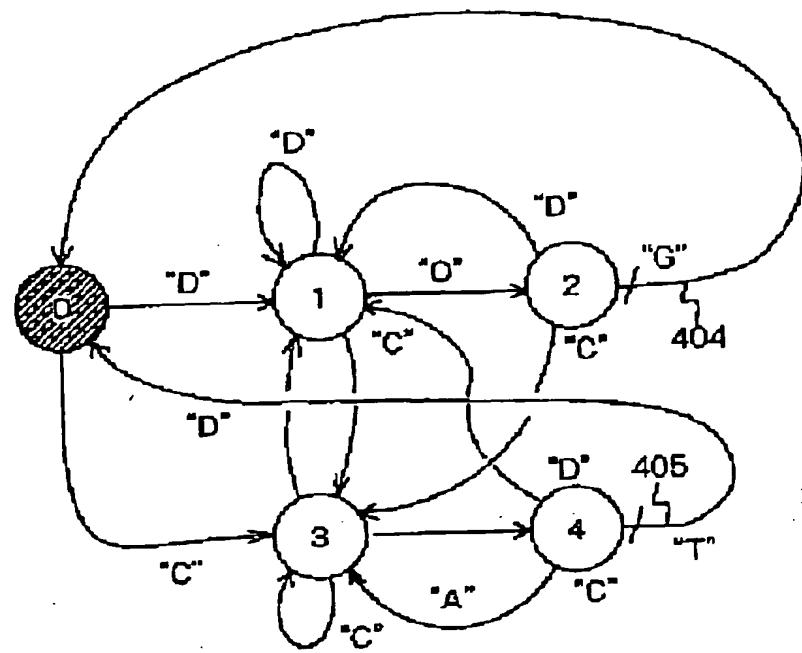


Fig. 4

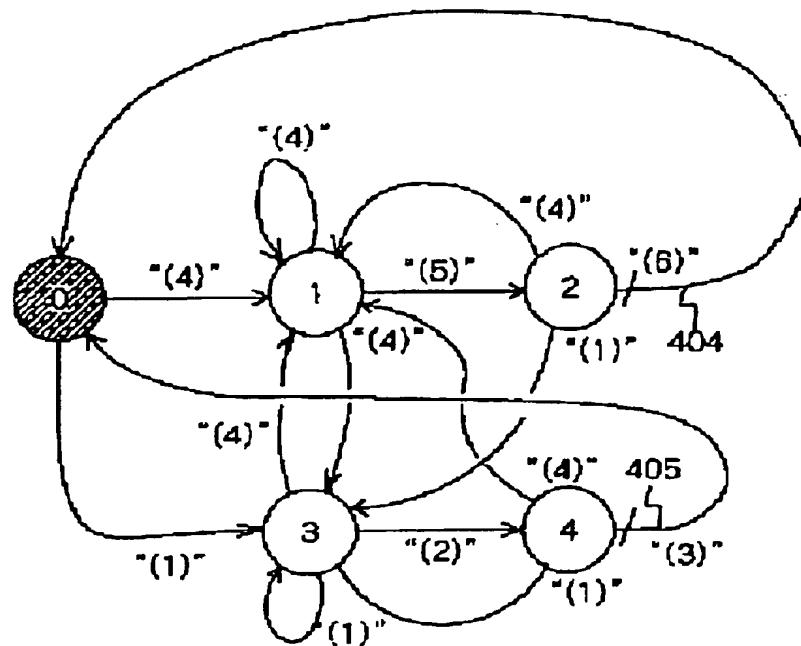


Fig. 8

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-139278

(43)公開日 平成6年(1994)5月20日

(51)Int.Cl.⁶
G 0 6 F 15/40

識別記号 庁内整理番号
5 0 0 E 7218-5L

F I

技術表示箇所

審査請求 未請求 請求項の数1(全10頁)

(21)出願番号 特願平4-287261

(22)出願日 平成4年(1992)10月26日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 川口 久光

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 多田 勝己

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 加藤 寛次

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 小川 勝男

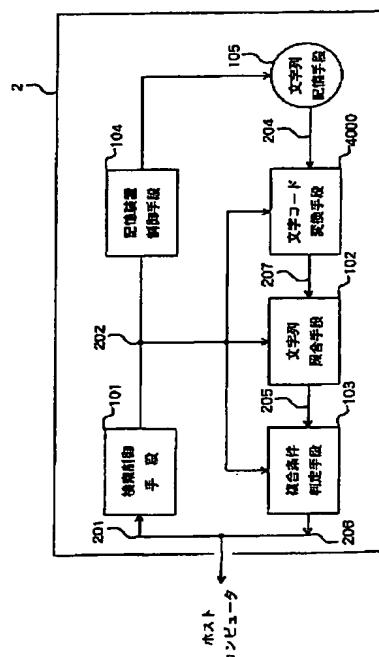
(54)【発明の名称】 文字コード変換機能を備えた文字列検索装置

(57)【要約】

【目的】 不要なスロットを削減することにより少量のメモリ量の状態遷移テーブルを実現し、コンパクトな文字列検索装置を提供することにある。

【構成】 テキストを格納する文字列記憶手段105と、文字列記憶手段から読み出されたテキストから1文字ずつ取り込み、文字コードが検索タームに含まれている文字コードの場合には文字に対応したコードを出力し、それ以外の場合には特定のコードを出力する文字コード変換手段4000と、予め与えられた検索タームの文字コードを対応するコードに変換する検索制御手段101と、文字コード変換手段の出力するコード列中に検索制御手段から送られた複数の検索タームに対応したコード列が存在するか否かを一括して照合する文字列照合手段102により文字列検索装置を構成する。

図1



【特許請求の範囲】

【請求項1】コード表現された文字で構成されるテキスト中に、指定された複数の検索タームが存在するか否かを一括して探索する文字列検索装置において、テキストを格納する文字列記憶手段と、該文字列記憶手段から読み出されたテキストから1文字ずつ取り込み、該文字コードが前記検索タームに含まれている文字コードか否かを判定し、検索タームに含まれている文字コードの場合には該文字に対応したコードを出力し、それ以外の場合には特定のコードを出力する文字コード変換手段と、予め与えられた検索タームの文字コードを対応する上記コードに変換する検索制御手段と、該文字コード変換手段の出力するコード列中に検索制御手段から送られた複数の検索タームに対応したコード列が存在するか否かを一括して照合する文字列照合手段と、を有することを特徴とする文字列検索装置。

【発明の詳細な説明】**【0001】**

【産業上の利用分野】本発明は情報処理システム、特に情報検索システムにおける検索タームの照合方法に係り、テキスト文字列から検索タームとして指定された複数の部分文字列が存在するか否かを一括して探索するためのものである。

【0002】データベース、文書ファイリングシステム、およびワードプロセッサなどにおける検索に利用し得るものである。

【0003】

【従来の技術】情報処理システムの分野では、文字列データの集まりからなる文書（以後、テキストと呼ぶ）の中から、検索者の探したいある特定の部分文字列（以後、検索タームと呼ぶ）を含む全ての文書を探し出すことが一つの重要な情報処理内容となっている。

【0004】このような検索システムを実現するための文字列検索装置がいくつか提案されている。その中の代表的な文字列検索装置の構成を図2に示し、その内容について説明する。

【0005】（エル エー ホラー：“テキスト リトリーバル コンピューターズ”、コンピューター、3月 1979年、L. A. Hollaar：“Text Retrieval Computers”，COMPUTER, March 1979）

文字列検索装置1において、検索制御手段101は、検索装置全体の制御と、ホストコンピュータとの通信を行う。すなわち、ホストコンピュータから送られてくる検索要求201を受け付け、これを解析し、文字列照合手段102と複合条件判定手段103へ検索情報202として送出する。また、検索制御手段101は記憶装置制御手段104を制御して、文字列記憶手段105に格納

された文字列データ204を文字列照合手段102へ読み出す。

【0006】文字列照合手段102は、入力文字列データ204の中に検索要求201に合致する文字列、すなわち検索タームがあるかどうかを調べ、もし該当するものががあれば、文字列を識別する情報205を複合条件判定手段103へ出力する。複合条件判定手段103は該文字列識別情報205に対して、検索要求201中に指示されたANDやORで構成される論理条件などが満足されるか否かを調べる。指定された複合条件を満足する場合には、該当する文書の識別情報や文書内容のテキストデータを検索結果206としてホストコンピュータへ返送する。

【0007】上述した文字列検索装置1の要となる文字列照合手段102における文字列の照合方式としては、有限オートマトンを用いて複数の文字列を1回のテキスト走査で探索する方法が知られている。その代表的なものとして、Ahoらが提案している方法がある。（エー・ブイ・エーホ アンド エム・ジェイ・コラッシック：“エフィシエント ストリング マッチング”、コミュニケーションズ エー シー エム、第18巻、第6号、1975年、A.V. Aho and M.J. Corasick：“Efficient String Matching”，CACM, VOL. 18, No. 6, 1975)。

【0008】そして、このAhoによるオートマトンを高速に実行するためのハードウェアとしては特開昭60-105039号公報により提案されている。ここで述べられている文字列照合回路について図3のブロック図を用いて説明する。

【0009】本従来技術は、文字コードレジスタ211、状態遷移テーブル220、状態番号レジスタ250、および照合IDテーブル260から構成される。

【0010】以下、本従来技術の文字列照合動作について説明する。

【0011】まず、初期設定として、状態遷移テーブル220には指定された検索タームを照合するためのオートマトンが設定される。さらにこのオートマトンの初期状態である状態番号0が状態番号レジスタ250に設定される。したがって、現状態番号305は状態番号0となる。

【0012】照合時動作は、入力文字列204から1文字ずつ文字コードレジスタ211に入力されることから始まる。文字コードレジスタ211の出力である文字コード302と状態番号レジスタ250の出力である現状態番号305をアドレスとして状態遷移テーブル220がアクセスされ、次に遷移すべき状態番号である次状態番号303が読み出される。次状態番号303は新たな現状態番号305として状態番号レジスタ250に保持される。

【0013】状態遷移テーブル220のアクセス動作に

並行して照合 1D テーブル 260 が状態番号レジスタ 250 から出力される現状態番号 305 と文字コード 302 をアドレスとしてアクセスされ、検索タームの識別番号が照合結果 205 として読み出される。ただし、読み出された検索タームの識別番号が 0 の場合、検索タームが照合されていないことを表わす。

【0014】その後、次の 1 文字が文字コードレジスタ 211 に入力される。

【0015】以上の一連の動作が繰り返されることにより文字列照合動作が実現される。

【0016】以下、本従来技術の動作について具体例で説明する。図 4 に示すオートマトンを例に用いる。

【0017】本図は、入力テキストの中から、検索者が与えた 2 つの検索ターム、すなわち “DOG” と “CAT” を照合するためのオートマトンの状態遷移図を示したものである。ここで、円はオートマトンの状態を、矢印は状態遷移を表す。各矢印に付記された文字はこれに対応した遷移を引き起こす入力文字を示している。各円の内部に記された数値は、その状態の状態番号を示す。状態 0 は、本オートマトンの初期状態である。遷移が記述されていない入力文字に対しては、全て初期状態 0 に遷移する。遷移にスラッシュ “/” が付記されている矢印 404 および矢印 405 はそれぞれ “DOG” および “CAT” が照合されたことを示す遷移を表す。すなわち、状態 2 から遷移する矢印 404 は “DOG” を、状態 4 から遷移する矢印 405 は “CAT” が照合されたことを示す。

【0018】以下、同図を用いて本従来技術の文字列照合動作について説明する。このオートマトンは初期状態 0 から状態遷移が始まる。初期状態 0 では、入力文字が、“D” であると状態 1 へ遷移し、“C” ならば状態 3 に遷移する。“D” および “C” 以外の文字が入力された場合は初期状態 0 に遷移する。状態 1 についても同様に、入力文字が “O” ならば状態 2 へ、“C” ならば状態 3 へ、“D” ならば状態 1 へ遷移し、それ以外は初期状態 0 へ戻る。状態 2 において、入力文字が “G” ならば照合結果が格納されている矢印 404 の遷移が起こり “DOG” が照合されたことになる。以下、他の状態における遷移についても同様である。

【0019】図 4 のオートマトンを格納した状態遷移テーブル 220 および照合 1D テーブル 260 の使用例をそれぞれ図 5 と図 6 に示す。

【0020】ここでは文字コードとして JIS コードを用いている。状態遷移テーブル 220 は入力文字コード 302 と着目しているオートマトンの現状態番号 305 でアドレッシングされる構成となっている。すなわち、現状態番号 305 が 0 で入力文字コード 302 が “D” のとき、0 と “D” に対応する 1 が次に遷移すべきオートマトンの次状態番号 303 として出力される。

【0021】照合 1D テーブル 260 には検索タームが

照合されたことを示す矢印 404 および矢印 405 の照合結果情報が格納されている。すなわち、検索タームの末尾文字で遷移するときの現状態番号と検索タームの末尾文字コード（図 4 の “DOG” の場合には状態 2 と文字 “G”）でアドレッシングされる照合 1D テーブル 260 の内容には検索タームの識別番号（以後、照合識別子とも呼ぶ）が格納されている。この “DOG” の例では、1 が照合識別子として格納されている。0 以外の内容が検索タームの識別番号を表している。照合識別子としては 0 以外の数値を割付け、照合対象以外には 0 を割付けることにより、照合出力を区別できるようにしている。

【0022】このようにして従来技術では検索処理が実行される。

【0023】

【発明が解決しようとする課題】状態遷移テーブルの全スロット数は状態数と文字コード数との積で表すことができる。本従来技術では、状態数として 256 状態、文字コード数として 256 種類用意しているため、スロット数は 64 K スロットとなる。

【0024】このように本従来技術では、検索タームにはわずかな文字コードの種類しか使用されないにもかかわらず、全ての文字コードに対応した状態遷移テーブルのスロットを用意しなければならない。図 4 のオートマトン例では 6 種類の文字コードしか使用されていない。さらに、図 4 のオートマトンに対応した図 5 の状態遷移テーブル例からも明らかなように、遷移情報が設定されているスロットは極わずかである。ここでは 1024 スロット中 12 スロット、すなわち約 100 分の 1 しか使用されていない。したがって、約 99% が不要なスロットとなっている。よって本従来技術には、状態遷移テーブルの使用効率が悪いため、不要に大きな状態遷移テーブルを必要とするという問題がある。

【0025】本発明の目的は、不要なスロットを削減することにより少量のメモリ量の状態遷移テーブルを実現し、コンパクトな文字列検索装置を提供することにある。

【0026】

【課題を解決するための手段】上記目的を達成するため、コード表現された文字で構成されるテキスト中に、指定された複数の検索タームが存在するか否かを一括して探索する文字列検索装置において、テキストを格納する文字列記憶手段と、該文字列記憶手段から読み出されたテキストから 1 文字ずつ取り込み、該文字コードが前記検索タームに含まれている文字コードか否かを判定し、検索タームに含まれている文字コードの場合には該文字に対応したコードを出力し、それ以外の場合には特定のコードを出力する文字コード変換手段と、予め与えられた検索タームの文字コードを対応する上記コードに変換する検索制御手段と、該文字コード変換手段の出力

するコード列中に検索制御手段から送られた複数の検索タームに対応したコード列が存在するか否かを一括して照合する文字列照合手段により文字列検索装置を構成する。

【0027】

【作用】上記手段を用いた本発明の照合原理について説明する。

【0028】まず、検索制御手段101において検索タームにどのような文字コードが指定されているかを調べ、これらの文字コードに例えばシリアル番号を割り付ける。このような文字コード対応テーブルを文字コード変換手段4000に設定する。さらにシリアル番号に変換した検索タームから作成したオートマトンを文字列照合手段102に設定する。

【0029】検索時には、文字列記憶手段102から読み出されたテキストが1文字ずつ文字コード変換手段4000に取り込まれ、該文字コードが前記検索タームに含まれている文字コードか否かが判定され、検索タームに含まれている文字コードの場合には該文字に対応したシリアル番号が出力され、それ以外の場合には特定の番号として例えば0(ゼロ)が出力される。

【0030】文字コード変換手段4000から出力されたコード列中に複数の検索タームに対応したコード列が存在するか否かが文字列照合手段102で一括して照合される。

【0031】以上述べた原理について具体例を用いて説明する。

【0032】まず、設定情報の作成方法について述べる。

【0033】本例では、検索タームとして図4のオートマトンと同様に“CAT”と“DOG”の2つを用いる。これらの検索タームには6種類の文字が使用されている。つまり、検索タームに使用されている文字の“C”、“A”、“T”、“D”、“O”、および“G”を、“(1)”、“(2)”、“(3)”、“(4)”、“(5)”、および“(6)”にそれぞれ対応させ、検索タームに使用されていない文字には“(0)”を対応させる。この文字コード変換テーブルが文字コード変換手段4000に設定される。このように設定された本例の文字コード変換テーブルを図7に示す。

【0034】また、検索タームも同様に“CAT”に対応する“(1)(2)(3)”と“DOG”に対応する“(4)(5)(6)”にそれぞれ変換される。このように変換された検索タームから作成されたオートマトンを図8に示す。図4のオートマトンの遷移に付記された文字が対応するシリアル番号に変換されている。このオートマトンが文字列照合手段102に設定される。

【0035】本例ではこのコード変換により、文字コードの種類が“(0)”～“(6)”の7種類に削減で

き、従来技術の256種類の約40分の1となる。したがって、状態遷移テーブルは約40分の1の容量で済むことになる。

【0036】次に照合動作について述べる。ここではテキスト例として“HOTDOG”を用いる。

【0037】まず、文字コード変換手段4000に“H”が入力されると図7に示す文字コード変換テーブルにより、対応する“(0)”が出力される。

【0038】次に、“O”が入力されると文字コード変換テーブルにより、対応する“(5)”が出力される。

【0039】同様に、“T”、“D”、“O”、“G”が次々に入力されると出力として“(3)”、“(4)”、“(5)”、“(6)”が出力される。つまり、出力テキストとして“(0)(5)(3)(4)(5)(6)”が文字列照合手段102に送出されることになる。

【0040】文字列照合手段102では文字コード変換手段4000から送出された出力テキスト中から“CAT”に対応する“(1)(2)(3)”または“DOG”に対応する“(4)(5)(6)”を図8に示すオートマトンに基づき探索する。ここでは、“(0)(5)(3)(4)(5)(6)”の入力により“DOG”に対応する“(4)(5)(6)”が照合される。

【0041】以上のように動作させることにより、状態遷移テーブルを従来技術の数十分の1のメモリ容量で実現でき、コンパクトな文字列検索装置を提供することが可能となる。

【0042】

【実施例】以下、本発明の原理を用いた第1の実施例について図1を用いて説明する。

【0043】本実施例では従来技術の文字列検索装置1の文字列記憶手段105と文字列照合手段102の間にコード変換手段4000を設けている。コード変換手段4000においては、文字列記憶手段から読み出した文字が検索タームに指定されている文字コードの場合には、この文字コードに対応した所定のコードに変換し、検索タームに指定されていない場合には特定のコードに変換し、文字列照合手段102に出力する。このようにすることにより文字列照合手段102で使用している状態遷移テーブルの文字コードに対応するエントリ数が検索タームに指定されている文字コード数+1となり、文字コードに対応したエントリ数を大幅に削減できるため少量のメモリで状態遷移テーブルが実現可能となる。

【0044】本実施例では、少量のメモリで状態遷移テーブルを実現することによりコンパクトな文字列検索装置を提供することを目的としている。

【0045】本実施例で用いる文字列検索装置の一構成例を図1に示す。文字コード変換4000は、文字列記憶手段105より読み出されたテキストから1文字入力される毎にこれを文字コードに対応した所定のコードに

変換し、文字列照合手段102へ出力する。

【0046】本文字コード変換回路は、文字コード変換テーブルから構成されている。

【0047】文字コード変換テーブル4100は図7に示す構成をとり、文字コードをアドレスとしてアクセスする一次元メモリである。本テーブルには、検索タームに使用されている文字コードの場合は、対応する予め定められたコードが設定され、その他の文字コードの場合は特定のコードが設定されている。

【0048】文字コード変換テーブル4100の具体例を用いて説明する。

【0049】ここでは、検索タームとして図4のオートマトンと同様に“CAT”と“DOG”の2つ検索タームを用いる。これらの検索タームには6種類の文字が使用されているため、検索タームに使用されている文字の“C”、“A”、“T”、“D”、“O”、および“G”を、“(1)”、“(2)”、“(3)”、“(4)”、“(5)”、および“(6)”にそれぞれ対応させ、検索タームに使用されていない文字には“(0)”を対応させる。このように設定した文字コード変換テーブルの具体例を図7に示す。

【0050】次にコード変換動作について述べる。ここではテキスト例として“HOTDOG”を用いる。

【0051】まず、コード変換手段4000に“H”が入力されると文字コード変換テーブルにより、対応する“(0)”が出力される。

【0052】次に、“O”が入力されると文字コード変換テーブルにより、対応する“(5)”が出力される。

【0053】同様に、“T”、“D”、“O”、“G”が次々に入力されると出力として“(3)”、“(4)”、“(5)”、“(6)”が出力される。つまり、出力テキストとして“(0)(5)(3)(4)(5)(6)”が文字列照合手段102に送出されることになる。

【0054】次に、文字列照合手段では、検索タームが“CAT”に対応する“(1)(2)(3)”と“DOG”に対応する“(4)(5)(6)”にそれぞれ変換され、これらより作成された図8に示すオートマトンが設定される。本オートマトンでは、図4のオートマトンの矢印に付記された文字が対応するコードに変換されている。

【0055】本例に対応した状態遷移テーブルを図9に示す。本テーブルではこのコード変換により、文字コード

の種類が“(0)～(6)”の7種類に削減でき、従来技術の256種類の約40分の1となっている。したがって、状態遷移テーブルは約40分の1の容量で済むことになる。

【0056】文字列照合手段102ではコード変換手段4000から送出された出力テキスト中から“CAT”に対応する“(1)(2)(3)”と“DOG”に対応する“(4)(5)(6)”を図7に示すオートマトンに基づき探索する。

【0057】ここでは、出力テキスト207として“(0)(5)(3)(4)(5)(6)”が文字列照合手段102へ入力され、“DOG”に対応する“(4)(5)(6)”が照合される。

【0058】以上説明したように本発明によれば、文字列記憶手段105と文字列照合手段102との間に文字コード変換手段を設けることにより、状態遷移テーブルを従来技術の数十分の1のメモリ容量で実現でき、コンパクトな文字列検索装置を提供することが可能となる。

【0059】

【発明の効果】本実施例によれば、小規模な文字コード変換手段を設けるだけで、状態遷移テーブルのサイズを数十分の1に縮小でき、コンパクトな文字列検索装置を提供することが可能となる。

【図面の簡単な説明】

【図1】本発明の文字列検索装置の説明図である。

【図2】従来の文字列検索装置の説明図である。

【図3】従来の文字列照合回路の構成を示すブロック図である。

【図4】従来のオートマトンの状態遷移図である。

【図5】従来の状態遷移テーブルの説明図である。

【図6】従来の照合IDテーブルの説明図である。

【図7】本発明を用いた文字コード変換テーブルの説明図である。

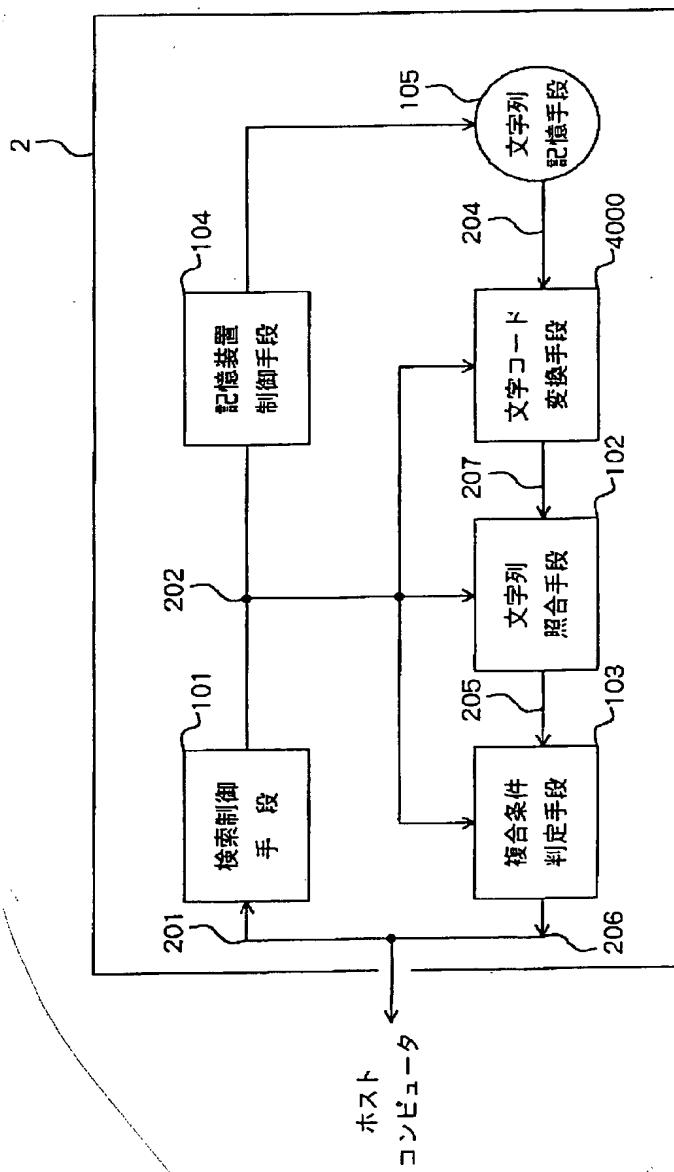
【図8】本発明を用いたオートマトンの状態遷移図である。

【符号の説明】

101…検索制御手段、102…文字列照合手段、103…複合条件判定手段、104…記憶装置制御手段、105…文字列記憶手段、220…状態遷移テーブル、260…照合IDテーブル、303…次状態番号、305…現状態番号、403…初期状態を示す矢印、4000…文字コード変換手段、4100…文字コード変換テーブル。

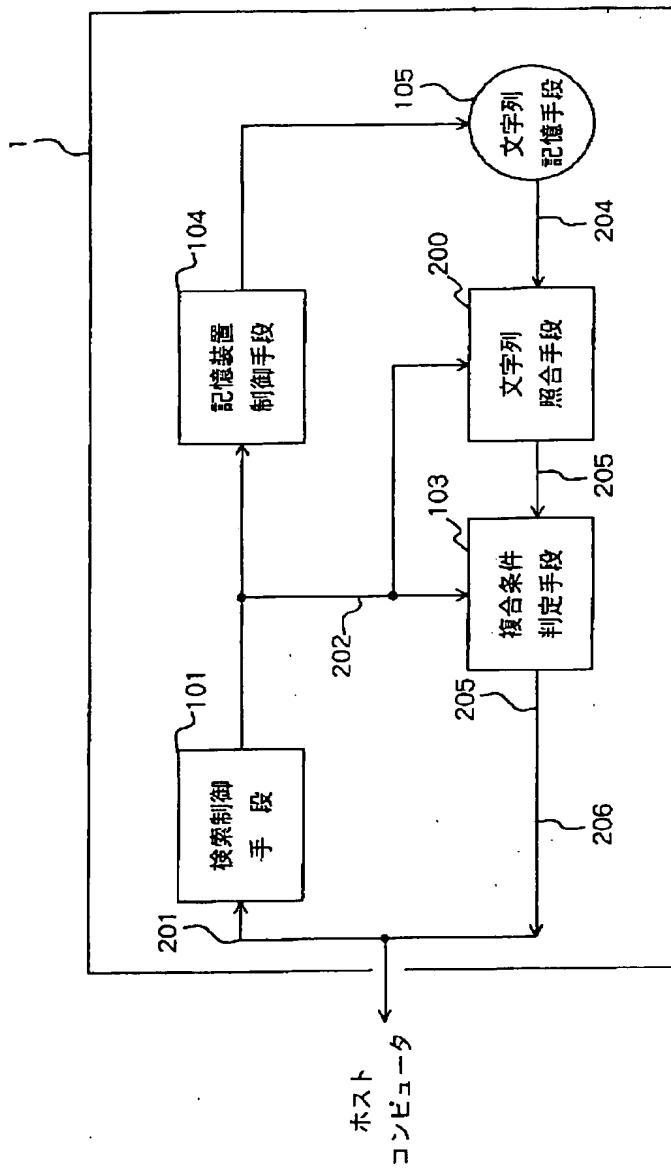
【図 1】

図 1



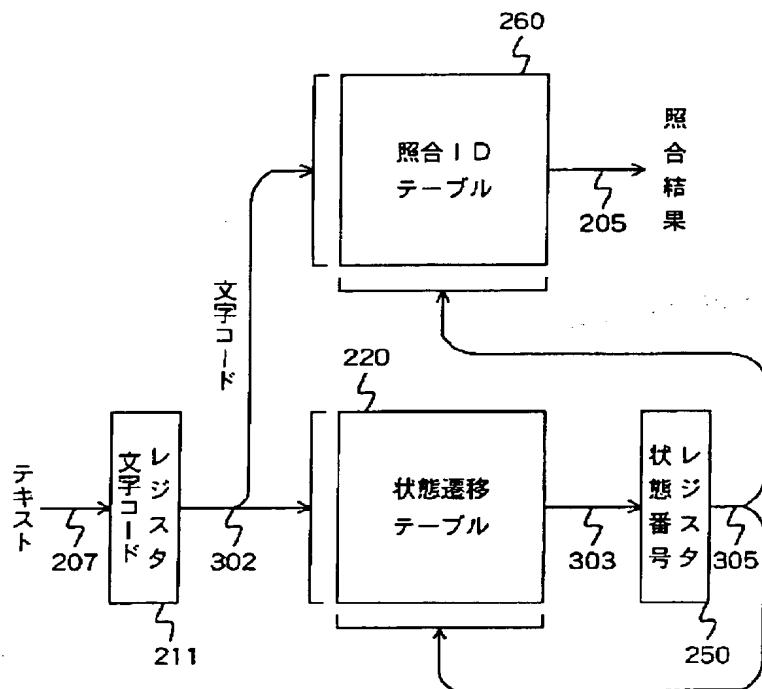
【図2】

図 2



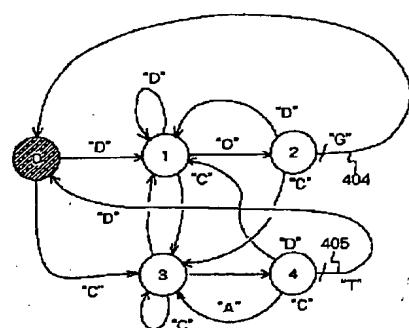
【図3】

図 3



【図4】

図 4



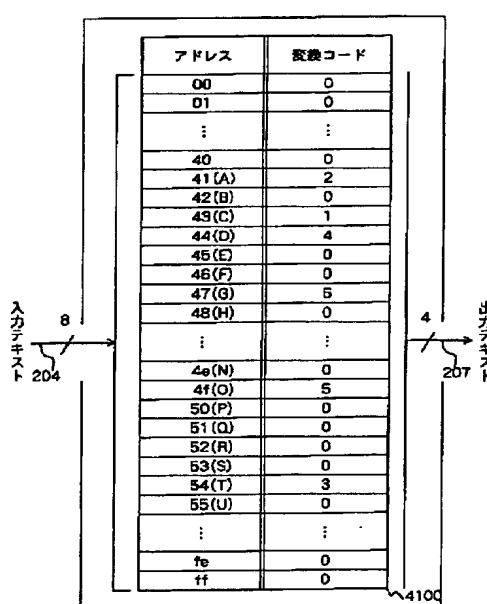
【図5】

文字			A	C	D	G	H	O	T									
文字	コード	0	1	…	41	…	43	44	…	47	48	…	4f	…	54	…	fe	ff
0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	3	1	0	0	0	0	2	0	0	0	0	0	0
2	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	4	0	3	1	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	4	0	3	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
fe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ff	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

("DOG" と "CAT" を照合するためのオートマトンに対応)

図 5

【図7】



4
207
出力テキスト
4100

【図6】

文字 コード				A	C	D	G	H	O	T								
状態番号		0	1	…	41	…	43	44	…	47	48	…	4f	…	54	…	fe	ff
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
.	
fe	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ff	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

【図9】

図9

文字コード		0	1	2	3	4	5	6	7
状態番号		0	1	2	3	4	5	6	7
0	0	3	0	0	4	0	0	0	0
1	0	3	0	0	4	2	0	0	0
2	0	3	0	0	4	0	0	0	0
3	0	9	2	0	4	0	0	0	0
4	0	3	0	0	4	0	0	0	0
5	0	0	0	0	0	0	0	0	0
.
fe	0	0	0	0	0	0	0	0	0
ff	0	0	0	0	0	0	0	0	0

【手続補正書】

【提出日】平成5年3月29日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】図面の簡単な説明

【補正方法】変更

【補正内容】

【図面の簡単な説明】

【図1】本発明の文字列検索装置の説明図である。

【図2】従来の文字列検索装置の説明図である。

【図3】従来の文字列使用号回路の構成を示すブロック図である。

【図4】従来のオートマトンの状態遷移図である。

【図5】従来の状態遷移テーブルの説明図である。

【図6】従来の照合IDテーブルの説明図である。

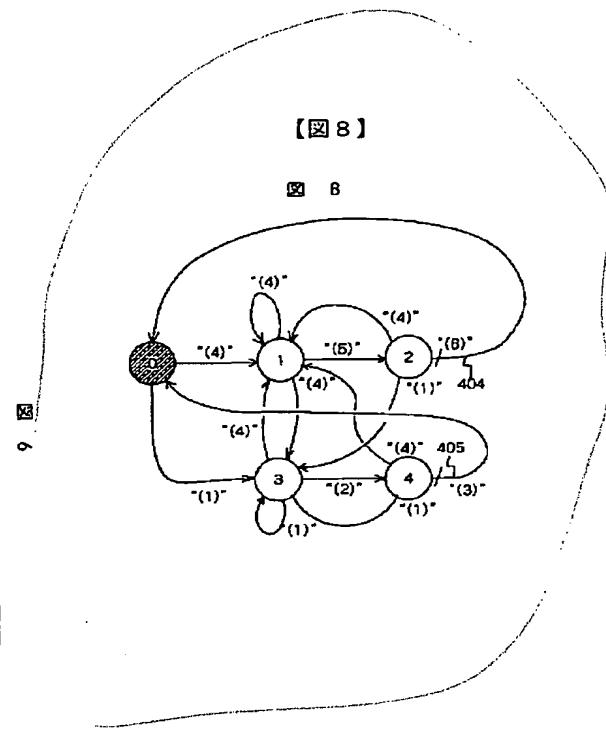
【図7】本発明を用いた文字コード変換テーブルの説明図である。

【図8】本発明を用いたオートマトンの状態遷移図である。

【図9】本発明による状態遷移テーブルの説明図である。

【符号の説明】

101…検索制御手段、102…文字列照合手段、103…複合条件判定手段、104…記憶装置制御手段、105…文字列記憶手段、220…状態遷移テーブル、260…照合IDテーブル、303…次状態番号、305…現状態番号、403…初期状態を示す矢印、4000



…文字コード変換手段、4100…文字コード変換テーブル。

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.